

# TFG Álvaro García Martín

*por* LUIS ANTONIO PUCH GONZALEZ

---

ARCHIVO 65224\_LUIS\_ANTONIO\_PUCH\_GONZALEZ\_TFG\_ALVARO\_GARCIA\_MARTIN\_7  
06051\_124980190.PDF (1.95M)

HORA DE LA ENTREGA 22-JUN.-2021 05:01P. M. (UTC+0200)

NÚMERO DE PALABRAS 7766

IDENTIFICADOR DE LA ENTREGA 1610657586

SUMA DE CARACTERES 38471



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID

Facultad de Ciencias Económicas y  
Empresariales

Grado en Economía

Trabajo de Fin de Grado

MODELIZACIÓN Y EVALUACIÓN DE POLÍTICAS  
CONTRA PANDEMIAS

**Autor:** Álvaro García Martín

**Tutor:** Luis Antonio Puch González

**Curso:** 2020-2021

**Convocatoria:** Junio

## Resumen

En este trabajo, profundizaremos en las principales herramientas con las que cuentan los Gobiernos a la hora de tomar decisiones para frenar la expansión de una pandemia. Para ello, se realiza un estudio de los principales modelos deterministas y estocásticos, se realizan simulaciones para cada uno de ellos y se propone el método de control sintético como modelo para evaluar las políticas adoptadas. Por otro lado, se estima el periodo en el que transcurre la primera ola en España y demás países europeos y se realiza un análisis descriptivo para observar sus características más representativas y tendencias. Finalmente, se realiza un análisis de control sintético con el fin de evaluar el primer estado de alarma en España como respuesta a la pandemia de COVID-19.

**Palabras Clave:** Pandemia, Control sintético, SIS, SIR, Cadenas de Markov.

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Revisión de la metodología</b>	<b>4</b>
2.1. Modelos deterministas: SIS y SIR . . . . .	4
2.2. Modelo SIR . . . . .	4
2.3. Modelo SIS . . . . .	6
2.4. Modelos estocásticos . . . . .	8
2.5. Cadenas de Markov . . . . .	8
2.6. Propagación de una epidemia usando cadenas de Markov . . . .	9
2.7. Control Sintético . . . . .	10
<b>3. Análisis descriptivo</b>	<b>11</b>
3.1. Datos . . . . .	11
3.2. Cálculo de la primera ola de la epidemia . . . . .	12
3.3. Análisis de los datos sin normalizar . . . . .	20
3.4. Análisis de los datos normalizados . . . . .	21
<b>4. Análisis de control sintético</b>	<b>22</b>
4.1. Resultados . . . . .	23
<b>5. Conclusión</b>	<b>26</b>
<b>6. Bibliografía</b>	<b>27</b>
<b>7. Anexo</b>	<b>28</b>

## 1. Introducción

El primer caso confirmado de COVID-19 en España, fue el 31 de enero de 2020. Poco después, empezaron a surgir casos relacionados con personas que regresaban de la región de Lombardía en Italia a lo largo del mes de febrero. El 13 de marzo de 2020, se habían confirmado casos de COVID-19 en las 50 provincias del país. La primera muerte a causa de COVID-19, ocurrió el 13 de febrero en Valencia, aunque no fue hasta el 3 de Marzo, cuando a través de una autopsia, se supo que estaba vinculada al COVID-19.

A nivel nacional, el número de casos y muertes aumentó rápidamente desde principio hasta finales de marzo, pasando de 2 muertes diarias el 1 de marzo a 870 el 31 de marzo. Ante la emergencia sanitaria, y en sintonía con otros países, el Gobierno español haciendo uso de artículo 116 de la Constitución y a través del Real Decreto 463/2020 de 14 de marzo, declaró el estado de alarma en todo territorio español, con el fin de garantizar la protección de la salud de los ciudadanos, la contención de la enfermedad y el refuerzo del sistema de salud pública.

Durante el periodo establecido se decretó la restricción de la circulación de los ciudadanos, la garantía en el suministro de alimentos y productos necesarios para la salud pública, la regulación de la apertura de la hostelería, la restauración y los locales donde se realizan actividades culturales, artísticas, deportivas y similares, la priorización del trabajo a distancia y la suspensión de la actividades escolares y educativas de manera presencial. El ministro de Sanidad tuvo bajos sus órdenes directas a las autoridades civiles de las Administraciones Públicas de todo el país, en particular las sanitarias, para asegurar su plena disposición y garantizar una mejor distribución territorial de medios técnicos y humanos.

El estado de alarma se estableció con una duración de 15 días, prorrogándose hasta seis veces y finalizando a las 00:00 horas del día 21 de junio de 2020 en las mismas condiciones establecidas en el mismo y en las disposiciones que lo modificaron, aplicaron y desarrollaron. El 28 de abril de 2020 se aprobó el Plan de Desescalada, en el cual se establecieron los principales parámetros e instrumentos para la adaptación del conjunto de la sociedad a la nueva normalidad.

Este trabajo tiene dos objetivos. El primero, explicar las diferentes herramientas con las que cuentan los Gobiernos y autoridades sanitarias, a la hora de desarrollar una estrategia de control frente a una epidemia. El segundo objetivo, evaluar el primer estado de alarma adoptado por el Gobierno español durante la primera ola de la epidemia.

Para ello, el trabajo se divide en tres secciones. Primero, se realiza una revisión metodológica de los principales modelos utilizados para estudiar las dinámicas de la enfermedades infecciosas. Analizamos el uso y aplicación de modelos deterministas y modelos estocásticos, y se propone el método de control sintético como modelo a la hora de evaluar las políticas adoptadas. En segundo lugar, calculamos el periodo en el que transcurre la primera ola de la pandemia para España y el resto de países y se realiza un análisis descriptivo del número de casos por millón de habitante para España y demás países europeos. Y por último, se realiza un análisis de control sintético para evaluar el estado de alarma en España; obteniendo resultados positivos a la hora de reducir el número de casos de COVID-19 en España.

## 2. Revisión de la metodología

La pandemia del COVID-19 ha evidenciado el elevado impacto que puede tener una epidemia a nivel global, tanto desde la perspectiva sanitaria como económica y social. Ante pandemias de estas características, los gobiernos y las distintas autoridades sanitarias deben actuar rápidamente e impulsar políticas de actuación que tengan como propósitos: limitar la transmisión del virus, evitar el colapso del sistema sanitario, reducir la mortalidad asociada al virus y, todo ello, condicionado por la necesidad de priorizar recursos en un contexto donde son escasos. En este sentido, el apoyo a la toma de decisiones mediante el uso de modelos matemáticos puede ser un punto clave. Explicar y predecir la velocidad y la forma de propagación del virus, apoyar la planificación sanitaria, identificar y estratificar el riesgo de los pacientes, así como evaluar la efectividad de las políticas tomadas son algunas de las aplicaciones potenciales de estas herramientas.

En esta sección explicaremos y discutiremos sobre el uso y aplicación de modelos deterministas y modelos estocásticos como herramientas clave para la toma de decisiones en situaciones de pandemia. Dentro de los modelos deterministas, hablaremos sobre los modelos SIS Y SIR. En cuanto a los modelos estocásticos profundizaremos en modelos basados en cadenas de Markov como herramienta complementaria para predecir el curso de una pandemia. Para la evaluación de las políticas adoptadas hablaremos sobre modelos de control sintético.

### 2.1. Modelos deterministas: SIS y SIR

En la sociedad actual, el contacto e interacción entre individuos juega un papel central. En el ámbito sanitario, tiene especial relevancia para determinar la propagación de una enfermedad. Uno de los modelos más utilizados sobre la propagación de una enfermedad a través del contacto e interacción de la gente es el modelo SIS (Susceptible, Infected, Susceptible). Siguiendo la explicación de **Jackson (2008)**, el modelo estudia la población total  $N$  dividida en dos nodos: ( $S$ )usceptibles e ( $I$ )nfectados, que evolucionan en un tiempo determinado  $t$ , las variaciones entre ellos son consecuencia del contagio por parte de individuos infectados a individuos susceptibles, e individuos infectados que se recuperan y vuelven a ser susceptibles. Esto supone una variación respecto al modelo SIR (Susceptible, Infected, Removed) desarrollado por **Kermack y McKendrick (1927)**, en el cual, los individuos que salen del nodo de los infectados no pueden ser infectados nuevamente, sino que terminan en el nodo de los ( $R$ )emoved; es decir, para el propósito del modelo, estos individuos se consideran permanentemente inmunes o fallecidos.

### 2.2. Modelo SIR

El modelo SIR, es un modelo determinista donde la población de estudio se divide en tres clases epidemiológicas:  $S(t)$ ,  $I(t)$  y  $R(t)$ . Tal y como elaboraron **Kermack y McKendrick (1927)**, el objetivo del modelo es medir la prevalencia, es decir, el número de casos de una enfermedad en la población en un tiempo dado. Este modelo se basa en ecuaciones diferenciales para describir la dinámica de los contagios en una población cerrada con  $N$  individuos, de manera que:

$$N = S(t) + I(t) + R(t) \quad (1)$$

La transmisión de la enfermedad se realiza a través del contacto entre individuos; un individuo infectado tendrá contacto con  $\beta N$  individuos en un tiempo determinado, donde  $\beta > 0$  es una constante que mide la velocidad de contacto. La probabilidad de contacto entre un infectado y un susceptible será por lo tanto  $S/N$ .

Por otro lado, los infectados se recuperan y abandonan el grupo de infectados a una razón proporcional al tamaño del grupo  $I(t)$ , con una constante  $\gamma > 0$ , denominada velocidad de recuperación. El modelo puede ser formulado con el siguiente sistema de ecuaciones diferenciales:

$$\begin{cases} \frac{dS}{dt} = -\beta S(t)I(t), & S(0) = S_0 = N - I_0 \\ \frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t), & I(0) = I_0 > 0 \\ \frac{dR}{dt} = \gamma I(t), & R(0) = 0 \end{cases} \quad (2)$$

Para ver la dinámica del modelo, es importante definir el número reproductivo básico del modelo  $R_0$ .  $R_0$  representa el número promedio de infecciones producidas por un individuo, durante su periodo infectado, en una población donde todos son susceptibles. Si este número es menor que uno, no habrá brote epidémico, ya que la incidencia es negativa, o lo que es lo mismo, la prevalencia es decreciente. Por el contrario, si  $R_0 > 1$ , entonces se desatará un brote epidémico que en su inicio tendrá un crecimiento exponencial. En el modelo SIR,  $R_0$  se define como:

$$R_0 = \frac{\beta}{\gamma} N \quad (3)$$

A continuación, podemos observar dos simulaciones para comprobar las dinámicas de las curvas del modelo cuando los parámetros toman distintos valores:

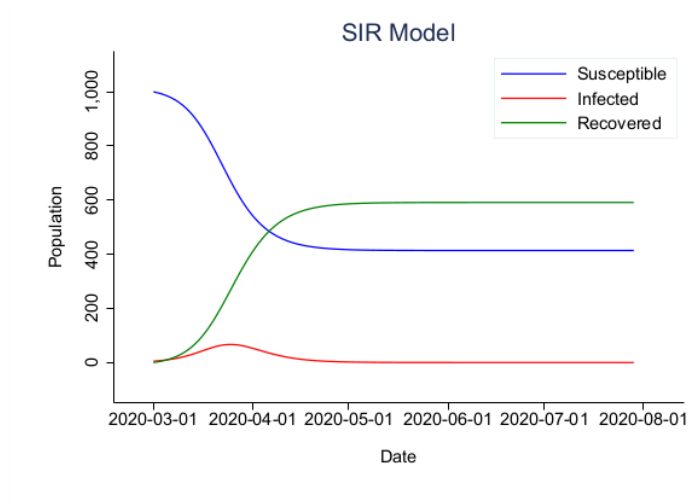


Figura 1: Dinámica de la epidemia con  $\beta = 0,5$ ,  $\gamma = 0,33$ ,  $I_0 = 5$  y  $N = 1000$

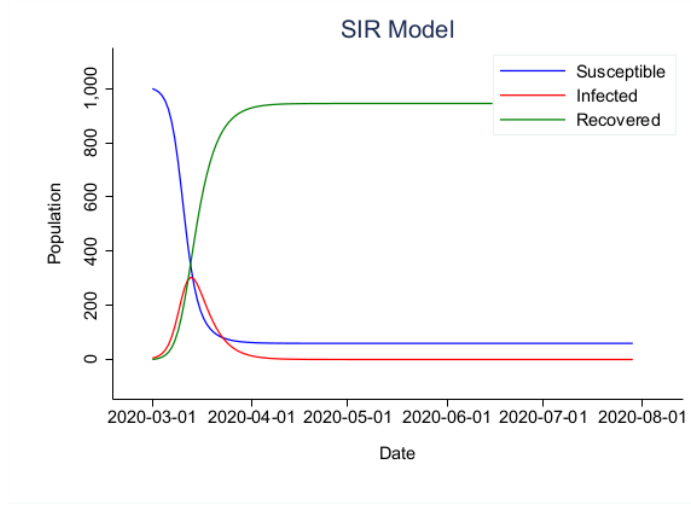


Figura 2: Dinámica de la epidemia con  $\beta = 0,75$ ,  $\gamma = 0,25$ ,  $I_0 = 5$  y  $N = 1000$

### 2.3. Modelo SIS

El modelo SIS es una variación del modelo SIR, en el que los infectados se recuperan y vuelven a ser susceptibles de infectarse de nuevo. Esto supone un avance a la hora de predecir enfermedades con riesgo de reinfección. El contagio se produce a través de una tasa de contagio  $\beta > 0$  que depende de cada enfermedad y de ambas poblaciones, mientras que la recuperación tiene lugar a través de una tasa de recuperación  $\gamma > 0$  que sólo depende de la población de individuos infectados que haya en cada momento. Por lo tanto, el modelo SIS puede ser formulado por el siguiente sistema de ecuaciones diferenciales:

$$\begin{cases} \frac{dS}{dt} = -\beta S(t)I(t) + \gamma I(t), & S(0) = S_0 > 0 \\ \frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t), & I(0) = I_0 > 0 \end{cases} \quad (4)$$

Al igual que en modelo SIR, en el modelo SIS, también es importante definir el número reproductivo básico para analizar la evolución de la enfermedad infecciosa. En el modelo SIS  $R_0$  toma la siguiente forma:

$$R_0 = \sigma = \frac{\beta N}{\gamma} \quad (5)$$

Para  $R_0 \leq 1$  la enfermedad se extingue. Para  $R_0 > 1$  la enfermedad permanece en la población de forma endémica. A continuación, podemos observar la dinámica del modelo con dos simulaciones para distintos valores de los parámetros.



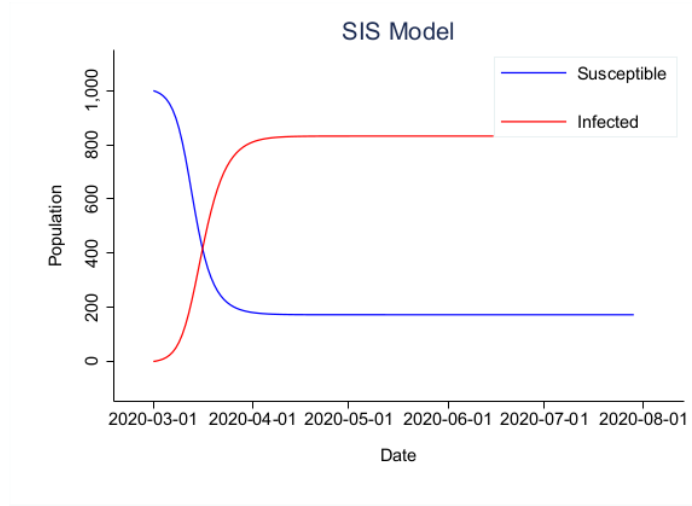


Figura 3: Dinámica de la epidemia con  $\beta = 0,2$ ,  $\gamma = 0,15$ ,  $I_0 = 5$  y  $N = 1000$



Figura 4: Dinámica de la epidemia con  $\beta = 0,01$ ,  $\gamma = 5$ ,  $I_0 = 5$  y  $N = 1000$

Los modelos deterministas vistos en esta sección, son modelos simples que ayudan a establecer posibles escenarios para el desarrollo de la enfermedad. Ante una emergencia sanitaria, la rapidez con la que se tomen las decisiones, puede ser crucial para el futuro desarrollo de la enfermedad. Ante este escenario, los modelos vistos en esta sección cuentan con la ventaja de proporcionar una ayuda inmediata mediante la estimación del  $R_0$ . No obstante, presentan una serie de desventajas, al aplicarse en tiempo real, los datos escasean y hay una gran incertidumbre sobre los parámetros. Esta gran incertidumbre y variabilidad en las estimaciones está causada en último término porque desconocemos el dato fundamental: cuántos pacientes hay realmente infectados. Solo sabemos cuántos han dado positivo, pero este dato depende del número de pruebas realizadas. Así pues, la variable endógena del modelo se mide con error.

## 2.4. Modelos estocásticos

En los modelos estocásticos, a diferencia de los modelos deterministas vistos anteriormente, los parámetros son variables aleatorias. En un modelo determinista, no se contempla el azar ni el principio de incertidumbre. En cambio, en los modelos estocásticos el azar sí interviene, y los posibles resultados se generan a través de la función de probabilidad. En el ámbito sanitario, los modelos estocásticos cumplen un papel fundamental a la hora de complementar los modelos deterministas, de manera que tanto en el modelo SIS como en el modelo SIR, se tome como hipótesis que los tiempos entre contactos o de recuperación sean variables aleatorias exponenciales, lo que da lugar a procesos markovianos.

## 2.5. Cadenas de Markov

Las cadenas de Markov son de una enorme utilidad en numerosos campos, en particular, en la predicción y el control del desarrollo de una epidemia. Una cadena de Markov es un proceso estocástico o conjunto de variables aleatorias (en cantidad discreta o continua, según el contexto) caracterizadas por la propiedad Markoviana; es decir, se refleja que el valor o estado de la variable asociada a un instante concreto de tiempo determina el siguiente estado del sistema en estudio, pero éste no depende de los estados asociados a las variables aleatorias anteriores. Para cualquier  $n > 0$  y para cualquier estado  $x_0, x_1, \dots, x_{n+1}$  se cumple:

$$P(x_{n+1} \mid x_0, x_1, \dots, x_n) = P(x_{n+1} \mid x_n) \quad (6)$$

En una cadena definida con  $m$  posibles estados  $E_1, E_2, \dots, E_m$ , se puede introducir la notación:

$$p_{ij} = P(x_n = j \mid x_{n-1} = i), \quad i, j = 1, \dots, m \quad (7)$$

para definir la probabilidad condicional de que estando en el estado  $i$  en el momento anterior inmediato pase al estado  $j$ .

- Si  $p_{ij} > 0$ , entonces el estado  $E_i$  está comunicado con el  $E_j$ . Si además  $p_{ji} > 0$ , la comunicación es mutua.

Los  $p_{ij}$  se denominan probabilidades de transición, y todas ellas forman la matriz de transición de dimensión  $m \times m$ .

$$T = [p_{ij}] \quad (8)$$

Dicha matriz, cumple las siguientes propiedades:

- Cada fila de la matriz es una distribución de probabilidad

$$\sum_{j=1}^m p_{ij} = 1 \quad (9)$$

- La potencia de la matriz de transición se puede tomar como:

$$T^n = CL^n C^{-1} \quad (10)$$

Donde  $L$  es la matriz diagonal con los valores propios de  $T$ , y  $C$  es la matriz que contiene los vectores propios asociados a los valores propios correspondientes.

## 2.6. Propagación de una epidemia usando cadenas de Markov

A continuación realizaremos una simulación para el modelo SIRM, de manera que la población se divide en los siguientes nodos: susceptibles, infectados, recuperados y fallecidos por la enfermedad. Asumimos el mes como unidad de tiempo  $n$ . Denotamos las probabilidades como:

- $\alpha$ : probabilidad de mantenerse en el estado de susceptible.
- $\beta$ : probabilidad de mantenerse en el estado de infectado.
- $\gamma = P(x_n = I \mid x_{n-1} = S)$ : probabilidad de estando en el estado susceptible, pasar al estado de infectado.
- $\mu = P(x_n = R \mid x_{n-1} = I)$ : probabilidad de estando en el estado de infectado, pasar al estado de recuperado.
- $\epsilon = P(x_n = M \mid x_{n-1} = I)$ : probabilidad de estando en el estado de infectado, pasar al estado de muerto.

La matriz de transición  $T$  de este proceso, tiene la siguiente forma:

$$\begin{pmatrix} \alpha & \gamma & 0 & 0 \\ 0 & \beta & \mu & \epsilon \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (11)$$

Otorgando valores a los parámetros de manera que  $\alpha = 0,35$   $\gamma = 0,65$   $\beta = 0,4$ ,  $\mu = 0,5$  y  $\epsilon = 0,1$ . Obtenemos la siguiente cadena de Markov:

**Cadena de Markov**

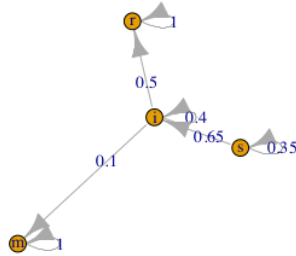


Figura 5: Cadena de Markov del modelo propuesto

A continuación se presenta una comparación entre los valores de las probabilidades de transición del estado susceptible al infectado, del estado muerte por enfermedad a partir del estado infectado y del estado de recuperado a partir del estado de infectado con el paso del tiempo:

Tiempo	$\gamma$	$\epsilon$	$\mu$
$n = 3$	0.274625	0.15600	0.56875
$n = 5$	0.06484156	0.1649600	0.7749219
$n = 7$	0.01293509	0.1663936	0.8220179
$n = 12$	0.000074739	0.166639	0.8331854

La probabilidad  $\gamma$ , a pesar de que va disminuyendo con el paso del tiempo, aún ocupa un lugar significativo en la dinámica de la epidemia y en la aparición de nuevos casos, provocando que crezca la probabilidad de muerte por la enfermedad  $\epsilon$ . Respecto a la probabilidad de recuperación de la enfermedad, va aumentando paulatinamente a lo largo del tiempo.

Esta herramienta matemática permite modelar y simular el comportamiento de la epidemia, gracias al análisis en el tiempo de los principales parámetros que caracterizan la dinámica de transmisión. El uso de cadenas de Markov a la hora de modelizar la epidemia, nos proporciona una serie de ventajas; nos permite considerar todos los orígenes, destinos y flujos de manera simultánea, así como la dependencia entre los flujos. Además, este tipo de análisis nos permite predecir sin conocer las causas de los fenómenos que modela. Es una herramienta fácil de usar en la que los cambios de sensibilidad se realizan de forma sencilla. Sin embargo, también cuenta con una serie de desventajas que son importantes de saber a la hora de predecir una epidemia. Las cadenas de Markov, suponen que la población es homogénea, lo que puede ser un problema ya que esto no refleja la realidad. También, Los cambios de estado ocurren en tiempo discreto, es decir, las transiciones tienen lugar en intervalos regulares de tiempo. Sin embargo, en algunos casos los cambios entre estados no ocurren regularmente, sino que puede ser el resultado de otras observaciones que tienen su propia distribución de probabilidad.

## 2.7. Control Sintético

Anteriormente, hemos visto modelos que servían como herramientas de apoyo a los gobiernos y autoridades sanitarias a la hora de diseñar la estrategia de control para frenar la expansión del virus. Una vez que la política adoptada está en marcha, es necesario realizar una evaluación de ella. En este trabajo, sugerimos el modelo de control sintético propuesto por **Abadie, Diamond y Hainmueller (2015)**. El objetivo del modelo, es estimar los efectos de políticas públicas que afectan a unidades agregadas como países o regiones. Para ello, comparan la evolución de variables de interés entre el país afectado por la intervención de la política pública y un grupo de control. El método de control sintético se basa en la observación de que una combinación de unidades no tratadas (es decir, un “control sintético”) frecuentemente proporciona una aproximación más cercana a las características de la unidad afectada por la intervención de la política pública que cualquier unidad individual. Emplea como unidad de control la media ponderada de las unidades no tratadas que mejor aproxima las características de la unidad tratada.

Suponiendo que hay  $J + 1$  países donde  $J = 1$  denota el país tratado (en nuestro caso sería España) y  $j = 2, \dots, J + 1$  son todos países no tratados. Se define a  $T_0$  como el momento del

tratamiento, en nuestro estudio es el 14 de marzo de 2020. Para España, conocemos los datos de contagios diarios confirmados ( $Y_{1t}$ ), pero se desconocen los contagios contrafactuales que se habrían producido si España no hubiera estado sujeta al tratamiento ( $Y_{1t}^N$  para  $t > T_0$ ). Por lo tanto, se tiene que encontrar una estimación de  $Y_{1t}^N$  para obtener una estimación del efecto de tratamiento  $\alpha_{it}$ :

$$\alpha_{it} = Y_{1t} - Y_{1t}^N \quad (12)$$

**Abadie et al (2015)** proponen hacer uso de las características observadas de los países en el grupo de donantes. Con la idea de encontrar pesos  $W = (w_2, \dots, w_{J+1})$  con  $w_j \geq 0$  para  $j = 2, \dots, J+1$  y  $\sum_{j=2}^{J+1} w_j = 1$ , de modo que el promedio ponderado de todos los países del grupo de donantes se asemeja al país tratado (España) con respecto al número de contagios en el período previo a la intervención y todos los demás aspectos relevantes. Formalmente, se busca  $W$  tal que:

$$\sum_{j=2}^{J+1} w_j Y_{tj} = Y_{1t} \quad \text{para} \quad \text{todo} \quad \sum_{j=2}^{J+1} w_j Z_j = Z_1 \quad (13)$$

Por lo que  $\sum_{j=2}^{J+1} w_j Y_{tj}$  para  $t \geq T_0$  es una estimación para los contagios contrafactuales no observados de  $Y_{1t}^N$  que induce a una estimación del efecto del tratamiento:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j Y_{tj}, \quad t \geq T_0 \quad (14)$$

Puede darse el caso de que un vector  $W$  de la ecuación (26) puede no existir (en particular, si existen diferencias estructurales entre España y los controles). Sin embargo, se pueden elegir los pesos tales como:

$$\min_w (X_1 - X_0 W)' V (X_1 - X_0 W) \quad (15)$$

Donde  $X_1$  denota un vector de  $k \times 1$  características previas a la intervención del país tratado, y  $X_0$  denota una matriz ( $k \times j$ ) de las mismas variables para los países  $j$  en el grupo de donantes.

### 3. Análisis descriptivo

En esta sección se realiza un análisis de la prevalencia del COVID-19 durante la primera ola en España, Portugal, Francia, Italia, Bélgica y Reino Unido, con la intención de de realizar una comparación entre todos los países, extraer sus características más representativas y describir sus tendencias. Ya que el virus afectó a los países en momentos distintos del tiempo, realizaremos el análisis con los datos normalizados. Además, también calcularemos el período que en que transcurre la primera ola a través del filtro Hodrick-Prescott.

#### 3.1. Datos

El análisis empírico se basa en datos diarios a nivel de país para el periodo 01/03/2020-31/07/2021. El motivo de elegir como variable dependiente el número de casos diarios por millón de habitantes, es evitar sesgos de mortalidad. Al inicio de una enfermedad nueva, los casos que primero se detectan son los casos más graves como los ingresos o las muertes, por lo

que es normal que la mortalidad de la enfermedad sea mayor a inicios de la pandemia. La base de datos utiliza para la realización del modelo, ha sido [Our World in Data](#), desarrollada en la Universidad de Oxford por el historiador social y economista Max Doser y su equipo. Estandariza y distribuye una de las bases de datos más extensa del mundo sobre salud, crecimiento, desigualdad, violencia, uso de energía, cultura y educación.

### 3.2. Cálculo de la primera ola de la epidemia

El periodo en el que transcurre la primera ola no ha sido definido por ningún gobierno o institución sanitaria. No hay un consenso científico para determinar los periodos que duran las olas, por lo que el criterio varía según el país, el medio o la institución que se consulte. En este trabajo utilizamos el filtro Hodrick-Prescott para precisar de un componente cíclico y un componente tendencial de la serie de estudio, con la finalidad de establecer un criterio con el que definir el periodo en el que transcurre la primera ola para los distintos países.

Siguiendo el trabajo desarrollado por **Hodrick y Prescott (1997)**, sea  $x_t$  la serie temporal de estudio,  $c_t$  el componente cíclico y  $\tau_t$  el componente tendencial; formalmente el filtro es definido como la solución del siguiente problema de optimización:

$$\text{Min} \sum_{t=1}^T c_t^2 + \lambda \sum_{t=2}^{T-1} (\nabla^2 \tau_t)^2 \quad (16)$$

Sujeto a:  $x_t = c_t + \tau_t$  y donde  $\nabla \equiv 1 - B$ ,  $B$  es el operador retardo:  $Bx_t = x_{t-1}$ .  $\lambda$  es el parámetro de suavidad de la tendencia extraída a la serie temporal. Si  $\lambda = 0$ , la solución degenera a  $\tau_t = x_t$  en cuyo caso el filtro excluye todas las frecuencias, es decir,  $c_t = 0$ . En el otro extremo, cuando  $\lambda \rightarrow \infty$  la solución se aproxima al ajuste lineal de mínimos cuadrados  $\tau_t = \beta_0 + \beta_1 t$ .

Para un  $\lambda$  fijo, se puede mostrar que el componente cíclico  $c_t$  puede ser calculado mediante:

$$c = (I_T - M^{-1})x \quad (17)$$

Donde  $I_T$  es la  $T \times T$  de la matriz identidad, y  $M$  es la matriz:

$$M = \begin{pmatrix} (1+\lambda) & -2\lambda & \lambda & 0 & 0 & 0 & \dots & 0 \\ -2\lambda & (1+5\lambda) & -4\lambda & \lambda & 0 & 0 & \dots & 0 \\ \lambda & -4\lambda & (1+6\lambda) & -4\lambda & \lambda & 0 & \dots & 0 \\ 0 & \lambda & -4\lambda & (1+6\lambda) & -4\lambda & \lambda & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \dots & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & \lambda & -4\lambda & (1+6\lambda) & -4\lambda & \lambda & 0 \\ 0 & \dots & 0 & \lambda & -4\lambda & (1+6\lambda) & -4\lambda & \lambda \\ 0 & \dots & 0 & 0 & \lambda & -4\lambda & (1+5\lambda) & -2\lambda \\ 0 & \dots & 0 & 0 & 0 & \lambda & -2\lambda & (1+\lambda) \end{pmatrix} \quad (18)$$

La serie que escogeremos para nuestro estudio son los datos diarios de los casos confirmados por millón de habitantes de COVID-19 en España. En la siguiente figura observamos la forma que tiene la serie:

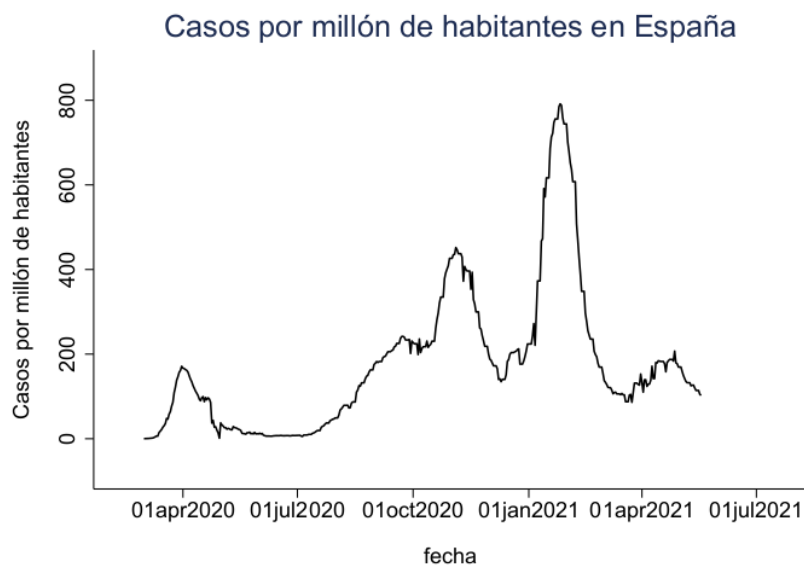


Figura 6: Elaboración propia a través de los datos de *Our World in Data*

En la serie, podemos apreciar diferentes picos que se han ido formando a lo largo del tiempo, alcanzado su máximo en enero de 2021. A partir de marzo de 2021, se aprecia un claro descenso del número de casos; probablemente causado por el efecto de la vacunación. Si nos centramos en la primera ola, es decir, el primer pico, parece que va desde principio de marzo hasta junio aproximadamente. Esta primera aproximación, nos sirve para darnos una primera idea del periodo en el que transcurre la primera ola.

Hay que tener en cuenta, que la serie está compuesta por datos diarios. Esto puede suponer un inconveniente, ya que el filtro Hodrick-Prescott está diseñado fundamentalmente para series con datos anuales, trimestrales y mensuales. Además, la mayoría de la literatura se enfoca en proporcionar el factor de suavidad  $\lambda$  óptimo para dichas series; por lo que encontrar un  $\lambda$  óptimo para series con datos diarios puede suponer un reto. No obstante, en este trabajo utilizaremos el factor de suavidad propuesto por **Weron y Zator (2015)**, los cuales utilizan un  $\lambda = 1600000$  para series con datos diarios.

Una vez conocido el factor de suavidad, procedemos a aplicar el filtro Hodrick-Prescott a nuestra serie de estudio, obteniendo los siguientes resultados:

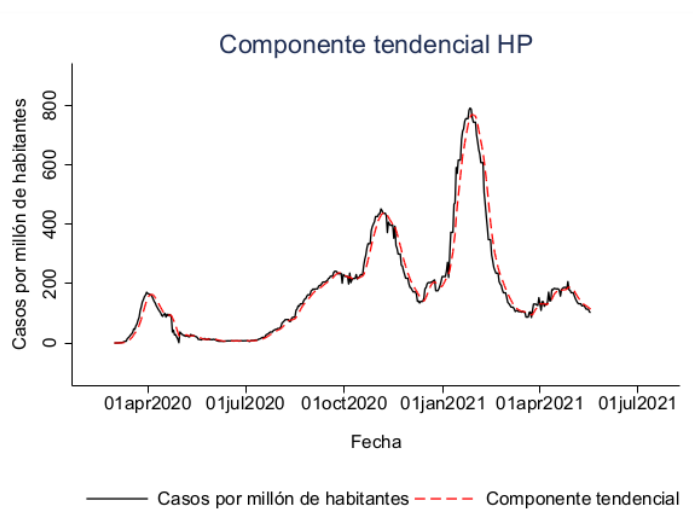


Figura 7: Elaboración propia a través de los datos de *Our World in Data*

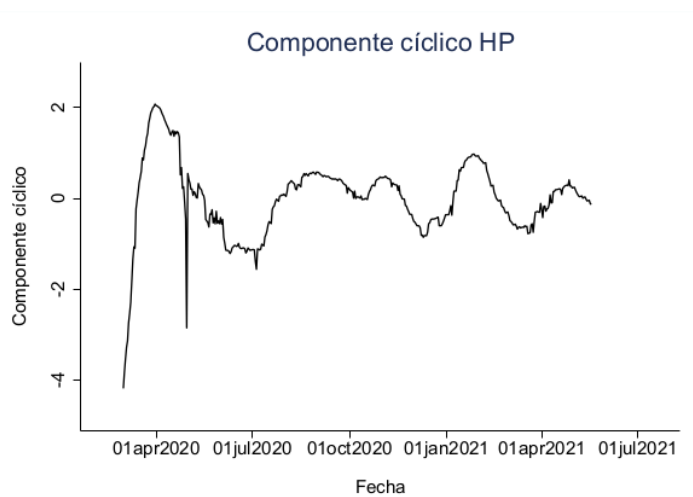


Figura 8: Elaboración propia a través de los datos de *Our World in Data*

Como podemos observar en la **Figura 8**, en el mes de mayo hay una abrupta caída debido a un atípico en la serie original. Para solucionar esto, lo interpolaremos calculando la media de los datos anteriores y posteriores de la caída antes de pasar el filtro. El nuevo componente tendencial tendría la siguiente forma:

En este nuevo gráfico se aprecia más claramente el periodo entre valle y valle, por lo que para España, podemos acotar la primera ola entre marzo y julio.



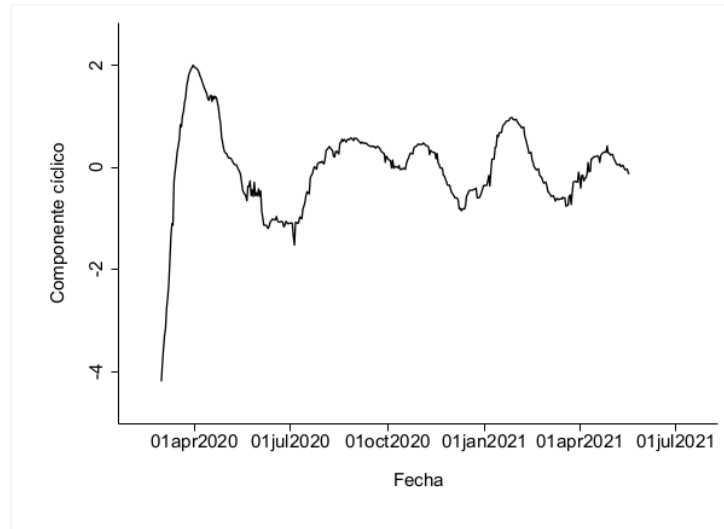


Figura 9: Elaboración propia a través de los datos de *Our World in Data*

A continuación, aplicaremos el filtro HP al resto de países con la finalidad de caracterizar su primera ola:

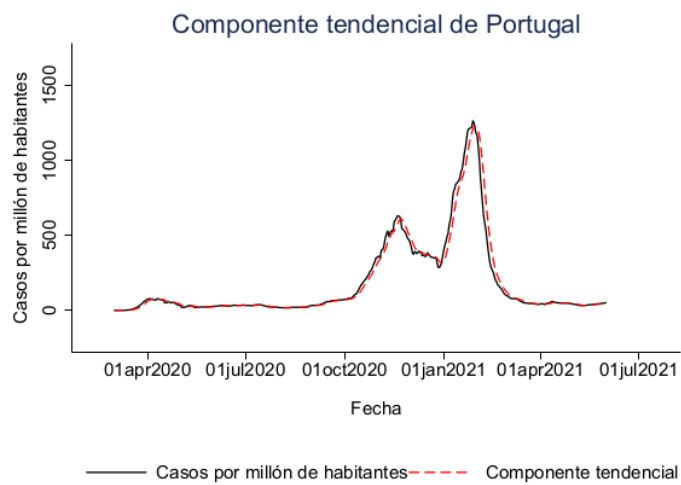


Figura 10: Elaboración propia a través de los datos de *Our World in Data*

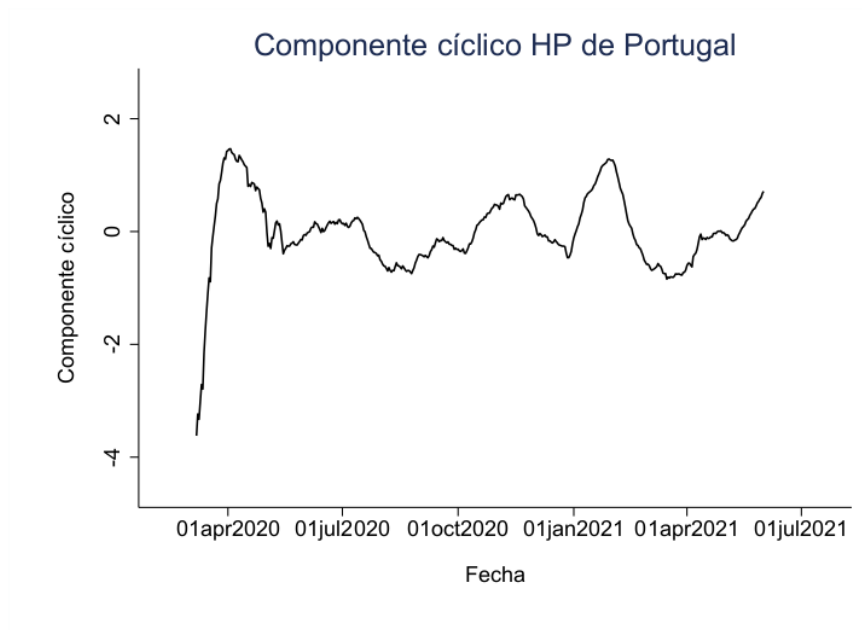


Figura 11: Elaboración propia a través de los datos de *Our World in Data*

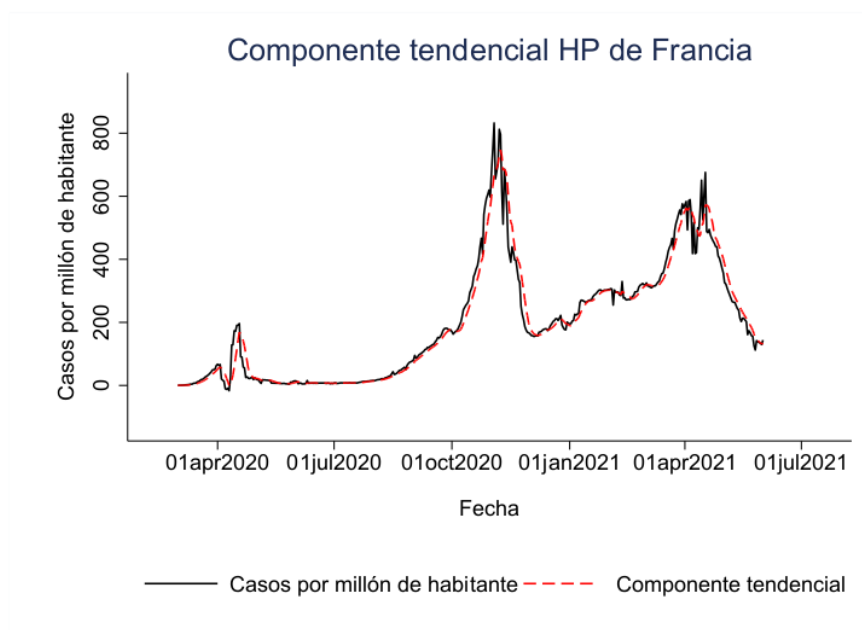


Figura 12: Elaboración propia a través de los datos de *Our World in Data*

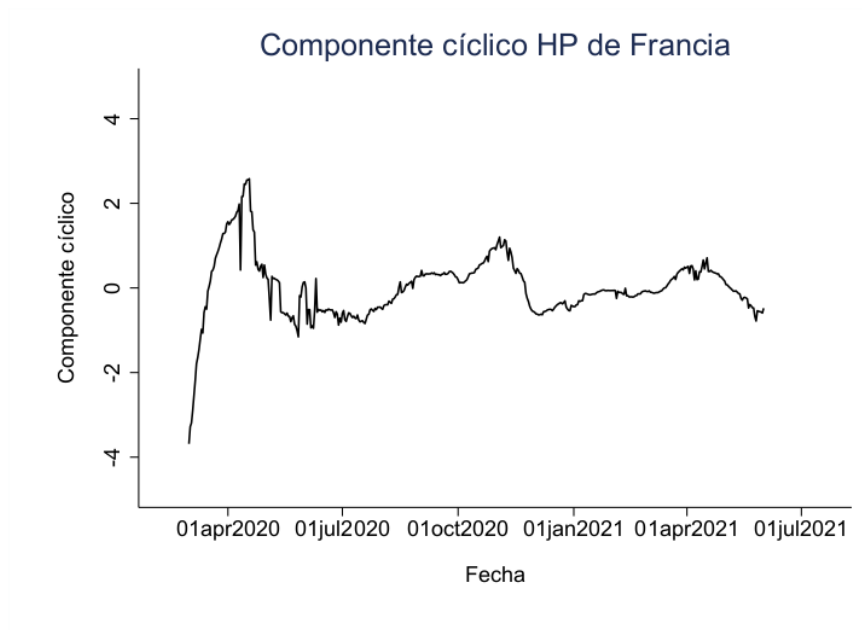


Figura 13: Elaboración propia a través de los datos de *Our World in Data*

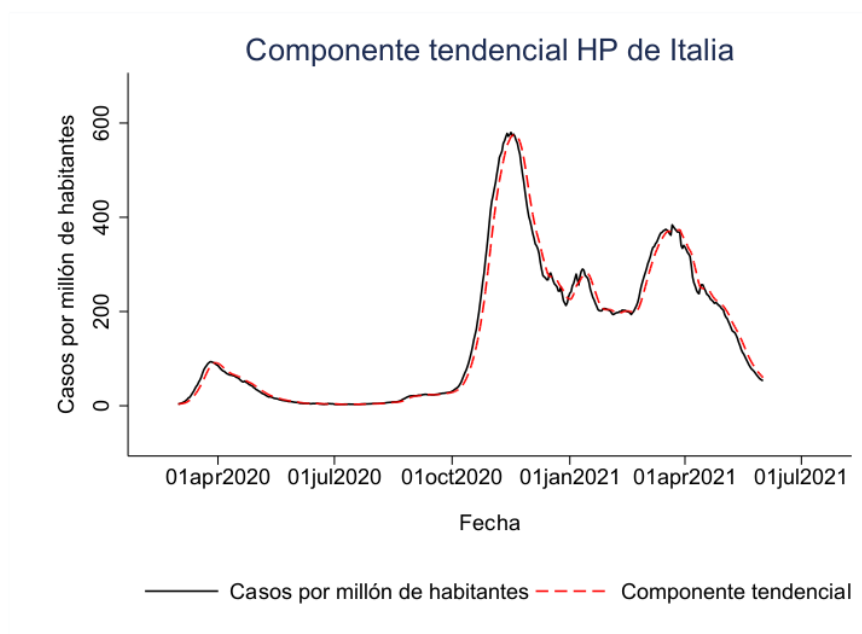


Figura 14: Elaboración propia a través de los datos de *Our World in Data*

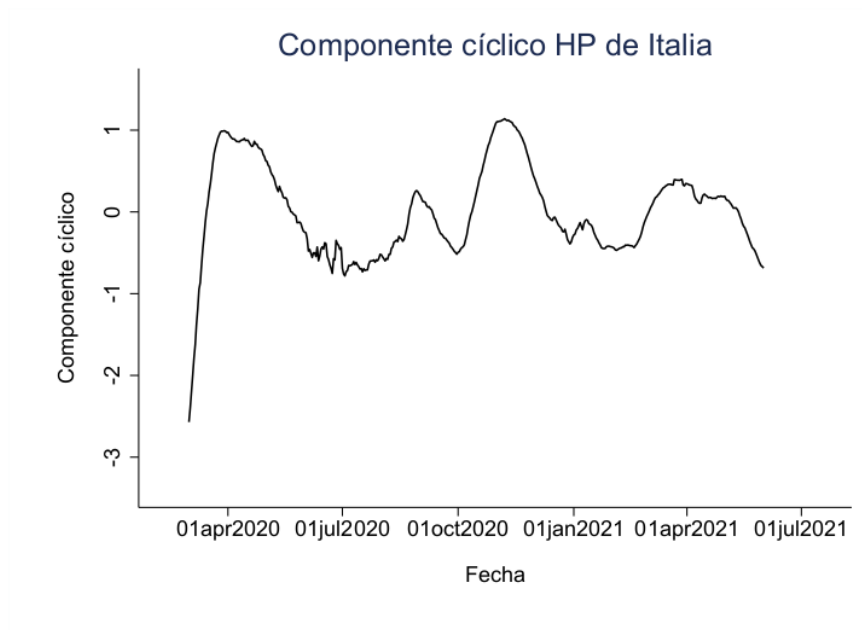


Figura 15: Elaboración propia a través de los datos de *Our World in Data*

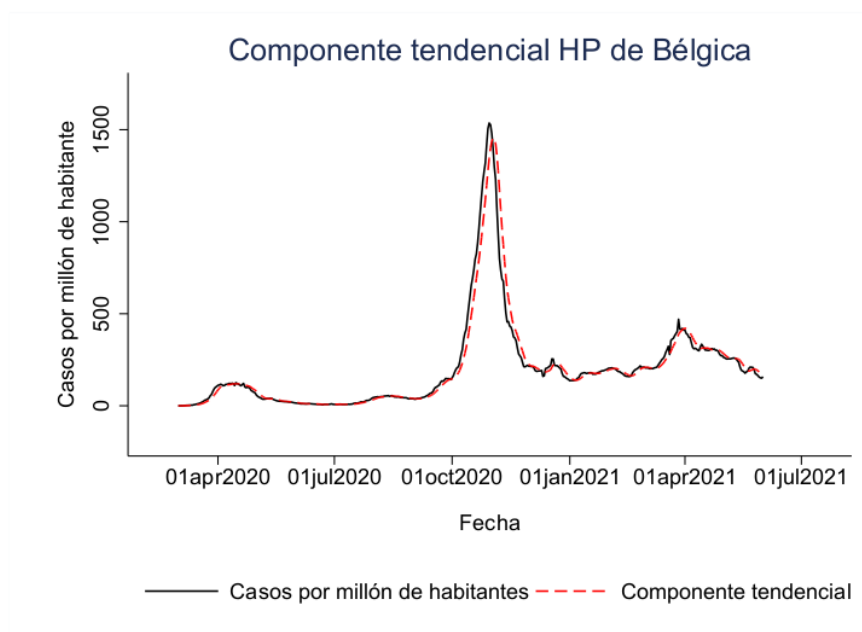


Figura 16: Elaboración propia a través de los datos de *Our World in Data*

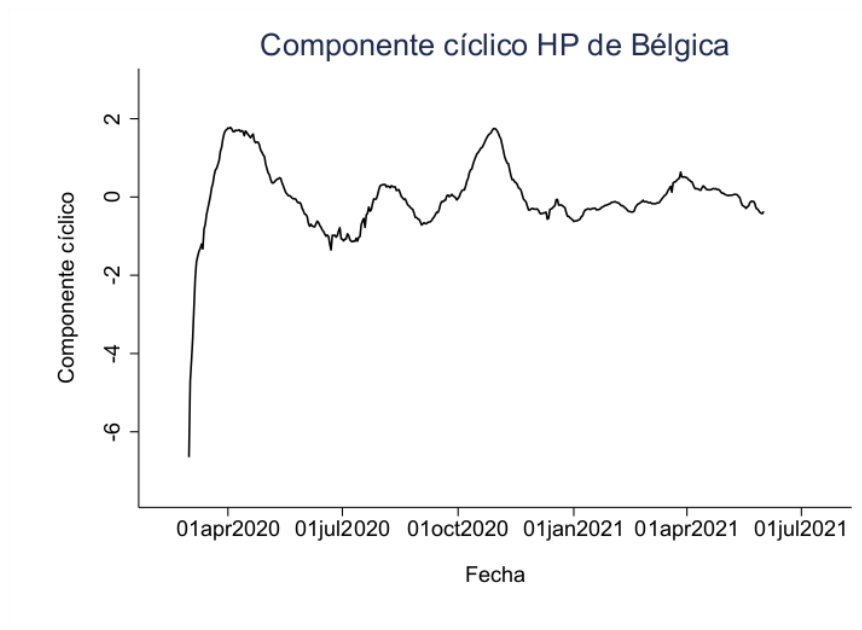


Figura 17: Elaboración propia a través de los datos de *Our World in Data*

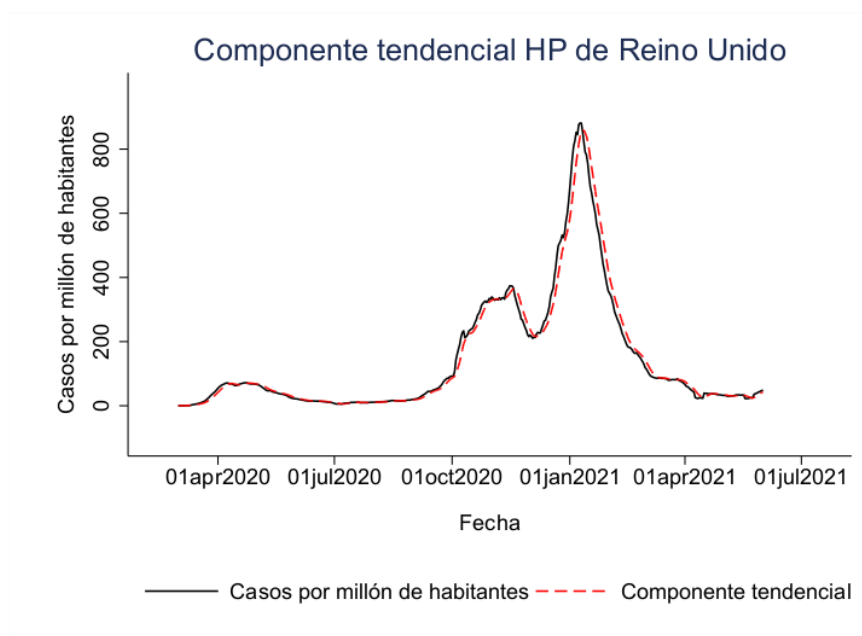


Figura 18: Elaboración propia a través de los datos de *Our World in Data*

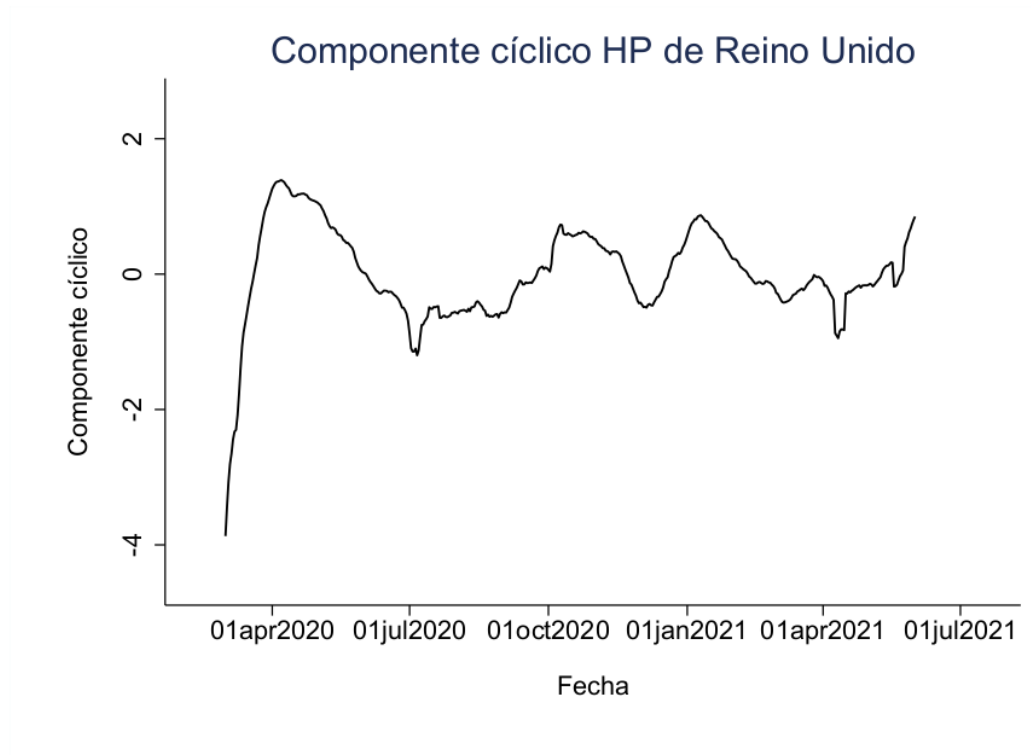


Figura 19: Elaboración propia a través de los datos de *Our World in Data*

La mayoría de países presentan un periodo similar entre valle y valle pudiendo acotar para la mayoría de ellos el periodo entre marzo y junio. En el caso de Francia, no está muy claro ya que hay varias abruptas caídas, como en el caso de España.

### 3.3. Análisis de los datos sin normalizar

Ahora que ya hemos definido el periodo en el que transcurre la primera ola, procederemos a analizar las tendencias y a comparar el impacto de la pandemia entre los países. En este primer enfoque, los datos estarán sin normalizar, una primera aproximación que nos servirá para ver la magnitud de la pandemia en cada país. Estudiaremos el número de casos por millón de habitantes para poder así analizar la prevalencia del virus en cada país.

Es interesante realizar este tipo de análisis ya que la prevalencia es un factor muy importante a la hora de determinar la estrategia o política adoptada para frenar la expansión del virus.

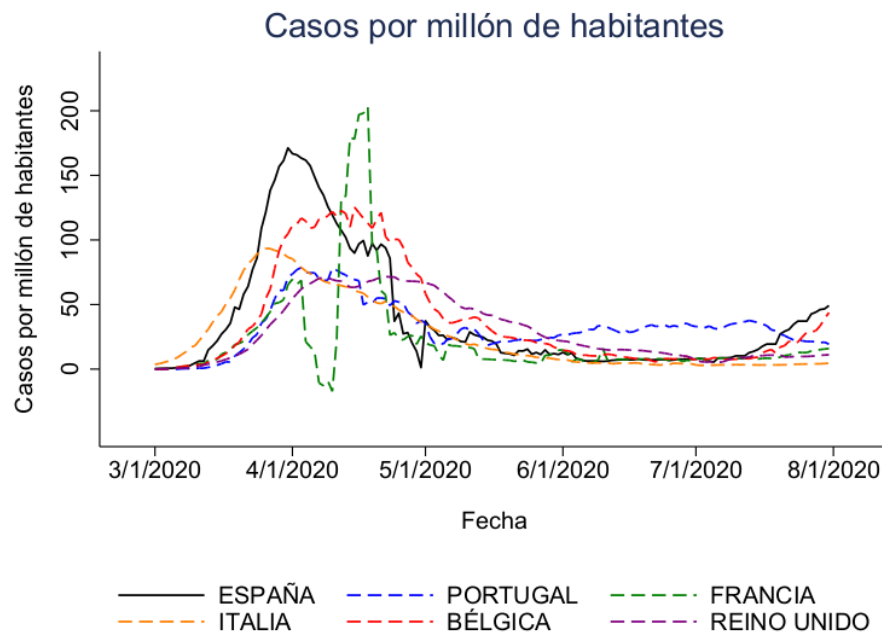


Figura 20: Elaboración propia a través de los datos de *Our World in Data*

La **Figura 20**, muestra la evolución del número de casos confirmados por millón de habitantes de COVID-19 por país durante la primera ola de la pandemia. Como podemos observar todos los países presentan una evolución similar, con un rápido incremento del número de casos, seguido de un descenso algo más lento. Italia es el país que primero destaca en el mes de marzo, ya que fue el primer país afectado por el virus; alcanza su pico a finales de marzo. Portugal y Reino Unido, se presentan como los países menos afectados durante la primera ola, teniendo ambos países una curva muy plana. Francia se muestra como el país más afectado durante la primera ola, presentando un primer pico a finales de marzo y un segundo pico, el más alto de la gráfica, a mediados de abril. A España la encontramos como el segundo país más afectado, teniendo un gran pico a finales de marzo. Para finalizar, destacamos la curva de Bélgica, distinta a las demás, presentando en el mes de abril un periodo de estabilización del número de casos, para luego comenzar el descenso.

### 3.4. Análisis de los datos normalizados

Como hemos visto en el apartado anterior, el virus afecta a los países en momentos distintos del tiempo. Por ello en este apartado normalizaremos los datos, esto nos proporcionará una comparación más exacta. Para normalizar los datos, hemos usado la incidencia acumulada a 14 días de cada país. La incidencia acumulada, se calcula dividiendo el número de casos confirmados en un periodo, entre el número de personas que están libres de la enfermedad al inicio del periodo. De manera que a través de una muestra estratificada se puede conocer la situación epidemiológica de una región o un país independientemente de cual sea su población total, equiparando así las diferencias entre territorios. En el caso del seguimiento diario del COVID-19 el criterio se ha fijado en una muestra de 100.000 habitantes. Los expertos señalan que el periodo que mejor refleja la situación epidemiológica de un territorio es la incidencia

acumulada de 14 días debido a los retrasos en las notificaciones de los nuevos contagios y los tiempos que tarda el virus en manifestarse en las personas. Los datos están ajustados de manera que el día 0, sea el día en el que cada país llega a una incidencia acumulada de 100.

A continuación, procederemos a analizar las tendencias y a comparar el impacto de la pandemia entre los países.

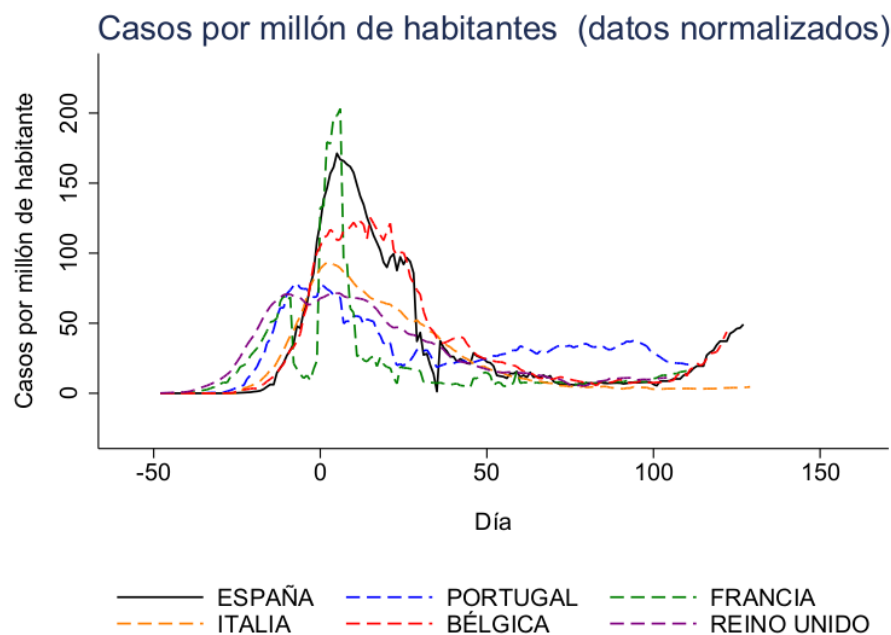


Figura 21: Elaboración propia a través de los datos de *Our World in Data*

Como podemos observar en la **Figura 21** las curvas de los países se ajustan al mismo periodo, facilitando la comparación entre ellos. Este proceso de normalización, es un paso necesario para el análisis de la **Sección 4**.

## 4. Análisis de control sintético

En la **Sección 2** vimos modelos que servían como herramientas de apoyo a los gobiernos y autoridades sanitarias a la hora de diseñar la estrategia de control para frenar la expansión del virus. Una vez que la política adoptada está en marcha, es necesario realizar una evaluación de ella. En esta sección realizaremos el análisis de control sintético propuesto en la **Sección 2** con el propósito de evaluar el primer estado de alarma en España y el plan de desescalada para la vuelta a la nueva normalidad, llevado a cabo por el gobierno español en 2020 como respuesta a la pandemia de COVID-19.

La fecha de inicio de la intervención es el 14/03/2020, día en el que se decretó el estado de alarma en España. Además, la base de datos utilizada está normalizada como vimos en la **Sección 3.3**. El grupo de países de control está compuesto por: Portugal, Francia, Italia, Bélgica y Reino Unido. Los países están seleccionados por sus similitudes con España, tanto culturales



como económicas y sanitarias. Nuestra variable dependiente son los contagios diarios por millón de habitantes, y nuestros predictores son: muertes por millón de habitantes, ingresados por millón de habitantes e ingresados UCI por millón de habitantes.

#### 4.1. Resultados

El **Cuadro 1** proporciona la lista de países de control y la participación de cada uno de ellos en la construcción de la España sintética. Como vimos en la **Sección 2.7**, los pesos están estimados de manera óptima, minimizando las diferencias entre España y la España sintética a partir del conjunto de datos. Para nuestra variable dependiente de estudio, el mejor contrafactual de España es un promedio ponderado de: Francia, Italia, Bélgica y Reino Unido. La proporción de Portugal es muy pequeña o cero, esto se debe a las diferencias en los datos entre Portugal y España durante la primera ola, como vimos en la **Sección 3.3**, la ola de Portugal es muy plana comparada con la de España. Francia tiene el peso más alto, seguido de Italia, ambos países con las olas más similares a la de España.

País	Unit Weight
Bélgica	0.18
Francia	0.409
Italia	0.408
Portugal	0.00
UK	0.003

Cuadro 1: Países de control y ponderación de cada uno en la España sintética

La pérdida, es decir, la raíz de la media cuadrática del error de predicción, captura la cantidad de error que hay entre los datos de España y los datos estimados de la España sintética. Se calcula mediante la siguiente fórmula:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (19)$$

Donde  $\hat{y}_i$  es el conjunto de datos estimados e  $y_i$  es el conjunto de datos conocidos. En nuestro análisis, la RMSE tiene un valor de :

<b>RMSE</b>	0.619227
-------------	----------

El **Cuadro 2** muestra el ajuste de los predictores para España y la España sintética. En general, los predictores se ajustan bastante bien, siendo los ingresados UCI el predictor que mejor se ajusta. El la **Figura 22** muestra la curva de contagios por millón de habitantes de España y la España sintética para el periodo de estudio. Como podemos comprobar, el estado de alarma tiene un efecto negativo en el número de contagios, haciendo que el pico de la primera ola de la pandemia sea menor en la curva de España respecto a la España sintética. Además, podemos apreciar que el pico de la España sintética es mucho más agudo, lo que se traduciría en un mayor número de casos en un tiempo menor; provocando un mayor colapso del sistema sanitario y como consecuencia un aumento de la letalidad del virus. Si analizamos la curva de España sintética observamos un primer pico, probablemente producido por el problema de las residencias en España que anticipa el segundo pico concentrado, (de Italia y Francia) después ambas curvas evolucionan a niveles similares hasta el comienzo de la segunda ola.

Predictores	España	España Sintética
Muertes por millón de habitantes	0.0431538	0.0341862
Ingresados por millón de habitantes	3.727462	3.273439
Ingresados UCI por millón de habitantes	0.676415	0.6797633

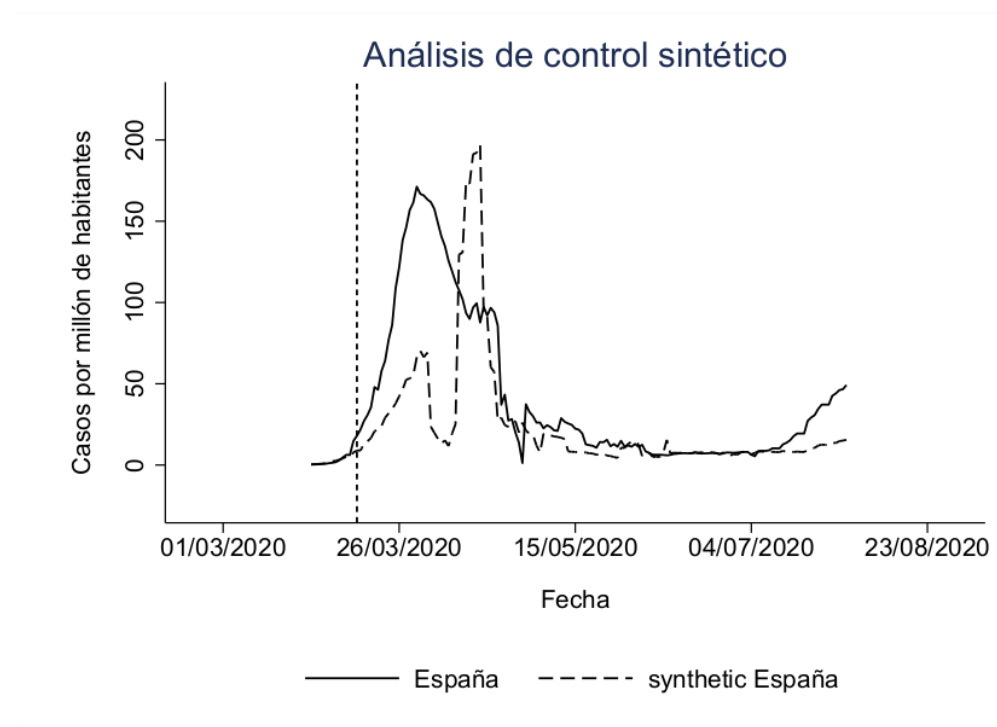


Figura 22: Elaboración propia a través de los datos de *Our World in Data*

Para hacer el análisis de control sintético, los pesos se han construido con datos del 1 al 14 de marzo para todos los países, y para construir la España sintética se utilizan los datos normalizados desde 100 casos sobre 100.000 habitantes. Como alternativa, realizaremos un segundo análisis utilizando datos de 0 a 100 casos sobre 100.000 habitantes para la construcción de los pesos. Como resultado obtenemos los siguientes resultados:

País	Unit Weight
Bélgica	0.00
Francia	0.829
Italia	0.171
Portugal	0.00
UK	0.00
<b>RMSPE</b>	0.478532

Predictores	España	España Sintética
Muertes por millón de habitantes	0.3683889	0.3684225
Ingresados por millón de habitantes	7.727462	6.984736
Ingresados UCI por millón de habitantes	1.552	1.84628

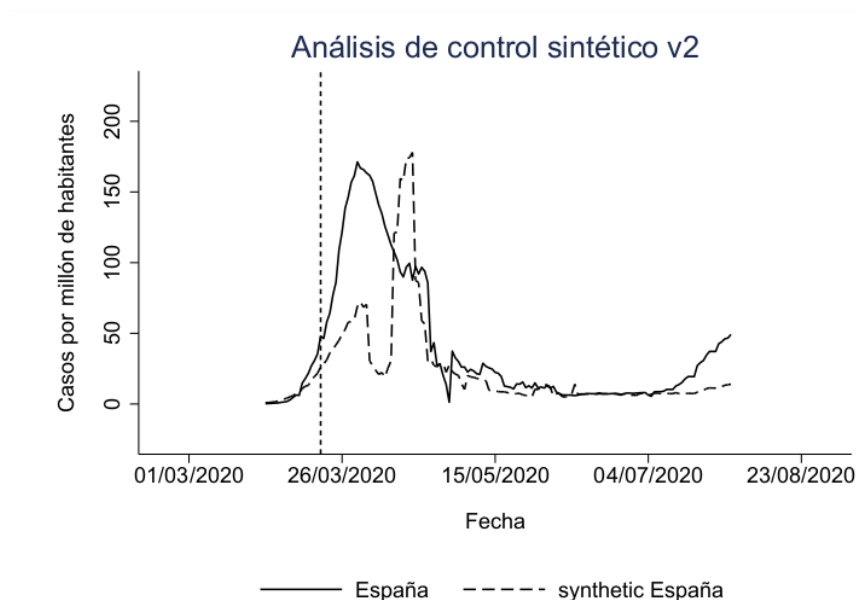


Figura 23: Elaboración propia a través de los datos de *Our World in Data*

Como podemos apreciar, los resultados obtenidos son muy similares a los del análisis inicial. Sin embargo, en este caso La España sintética está construida a través de los pesos ponderados de Francia e Italia; debido a que son los países con la incidencia más parecida a España al comienzo de la primera ola. La RMSPE es más baja que en el primer análisis por lo que el error cometido al estimar los datos de la España sintética es menor.

## 5. Conclusión

En este trabajo se han estudiado algunos modelos matemáticos para estudiar el comportamiento de la propagación de enfermedades infecciosas. Dichos modelos, ayudan a predecir y a preparar con antelación los medios necesarios para disminuir los efectos de la epidemia; estudiando los distintos escenarios posibles. Estos modelos, se dividen en 2 grandes grupos: modelos deterministas y modelos estocásticos. La realización de simulaciones es de gran utilidad para la toma de decisiones, para la mejora de la prevención y de programas de control, ya que ayudan a determinar qué decisión es la más acertada según el escenario en el que nos encontremos. Sin embargo, estos modelos presentan inexactitudes, una crítica hacia la forma de ajustar los modelos de propagación de epidemias es que de manera frecuente se ajustan los parámetros de acuerdo con las estadísticas de infectados proporcionadas por los gobiernos. Sin embargo, en pocos países existe una evidencia clara de que estas cifras reflejen la realidad debido al desconocimiento acerca del porcentaje de pacientes asintomáticos y la falta de realización de test de forma general entre la población.

Es por ello, que es imprescindible una constante evaluación de la política adoptada. En nuestro trabajo optamos por el método de control sintético como herramienta para evaluar la política pública. En el caso del COVID-19, el Gobierno español, se enfrentó a una crisis sanitaria nunca vivida antes, decretando el estado de alarma. La evaluación realizada en este trabajo obtiene resultados positivos respecto al estado de alarma. Logrando una disminución del número de contagios y su consecuente disminución en la tasa de letalidad.

Como líneas futuras de investigación, sería interesante el desarrollo de modelos discretos que proporcionen y completen una nueva visión sobre las dinámicas de las enfermedades infecciosas. Así como modelos de evaluación que consideren el ámbito sanitario y el económico, para así poder conocer las consecuencias económicas y sanitarias del plan de control.

## 6. Bibliografía

- Abadie, A., Diamond, A. y Hainmueller, J. (2015). "Comparative politics and the synthetic control method". *American Journal of Political Science*, 59, 495-510.
- Abadie, A., Diamond, A. y Hainmueller, J. (2010). "Synthetic control methods for comparative case studies: estimating the effect of California's Tobacco Control Program". *Journal of the American Statistical Association*, 105, 493-505.
- Abadie, A. y Gardeazabal, J. (2003). "The economic costs of conflict: a case study of the Basque Country". *American Economic Review*, 93, 113-132.
- Aleman, C., Busch, C., Ludwig, A., Santaaulalia-Llopis, R. (2020). "Evaluating the Effectiveness of Policies Against a Pandemic". *Barcelona Graduate School of Economics*.
- Braun, M. (1983). "Differential equations and their applications". *Applied Mathematical Sciences, Springer, New York*.
- Delgado Moya E.M (2017)., "Modelos para la Transmisión del VIH/SIDA y el Dengue. Estrategia de control". *Tesis en opción al título de Máster en Ciencias Matemáticas, Universidad de La Habana, la Habana, 2016, 73 p.*
- García, L. (2017). "Modelos matemáticos compartimentales para la epidemiología". *Universidad la Laguna*.
- Hodrick, R., y Prescott, E. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1), 1-16.
- Jackson, M. (2010). "Social and Economic Networks". *Princeton University Press, Princeton*, cap 7, 254-273.
- Kermack, W.O. McKendrick, A.G. (1927). "A contribution to the mathematical theory of epidemics". *Proceedings of the Royal Society* 115, 700-721.
- M. Matabuena, C. Mejjide-García, P.Rodríguez-Mier y V. Leborán (2020). "COVID-19: Estimating spread in Spain solving an inverse problem with a probabilistic model". *Cornell University Journal*
- Weron, R. y Zator M. (2015). "A note on using the Hodrick-Prescott filter in electricity markets". *Energy Economics*, 44, 178-190.

## 7. Anexo

### Simulación modelo SIR:

Population	2020-03-01	2020-03-25	2020-07-20
Susceptible	1000	672	414
Infected	5	67	0
Recovered	0	266	591
Total	1005	1005	1005

Cuadro 2: Modelo 1:  $\beta = 0,5$ ,  $\gamma = 0,33$ ,  $I_0 = 5$  y  $N = 1000$

Population	2020-03-01	2020-03-13	2020-07-20
Susceptible	1000	342	59
Infected	5	304	0
Recovered	0	359	946
Total	1005	1005	1005

Cuadro 3: Modelo 2:  $\beta = 0,75$ ,  $\gamma = 0,25$ ,  $I_0 = 5$  y  $N = 1000$

### Simulación modelo SIS:

Population	2020-03-01	2020-03-09	2020-05-20
Susceptible	1000	450	205
Infected	5	555	800
Total	1005	1005	1005

Cuadro 4: Modelo 1:  $\beta = 0,2$ ,  $\gamma = 0,15$ ,  $I_0 = 5$  y  $N = 1000$

Population	2020-03-01	2020-04-01	2020-07-29
Susceptible	1000	893	504
Infected	5	112	501
Total	1005	1005	1005

Cuadro 5: Modelo 2:  $\beta = 0,01$ ,  $\gamma = 5$ ,  $I_0 = 5$  y  $N = 1000$

## Test de placebo

Para evaluar la capacidad del método de control sintético para reproducir la evolución de la España sintética, **Abadie y Gardeazabal (2003)**, introducen un estudio de placebo. La idea del test de placebo propuesto, es evaluar la significación de nuestras estimaciones.

Si el test de placebo crea brechas de magnitudes similares al estimado para España, entonces nuestra interpretación es que nuestro análisis no proporciona evidencia significativa del efecto negativo en el número de casos provocado por el estado de alarma. Si por el contrario, el test de placebo muestra que la brecha estimada para España es inusualmente grande en relación con la brecha del resto de países, entonces nuestra interpretación es que nuestro análisis proporciona evidencia significativa de un efecto negativo en el número de casos debido al estado de alarma.

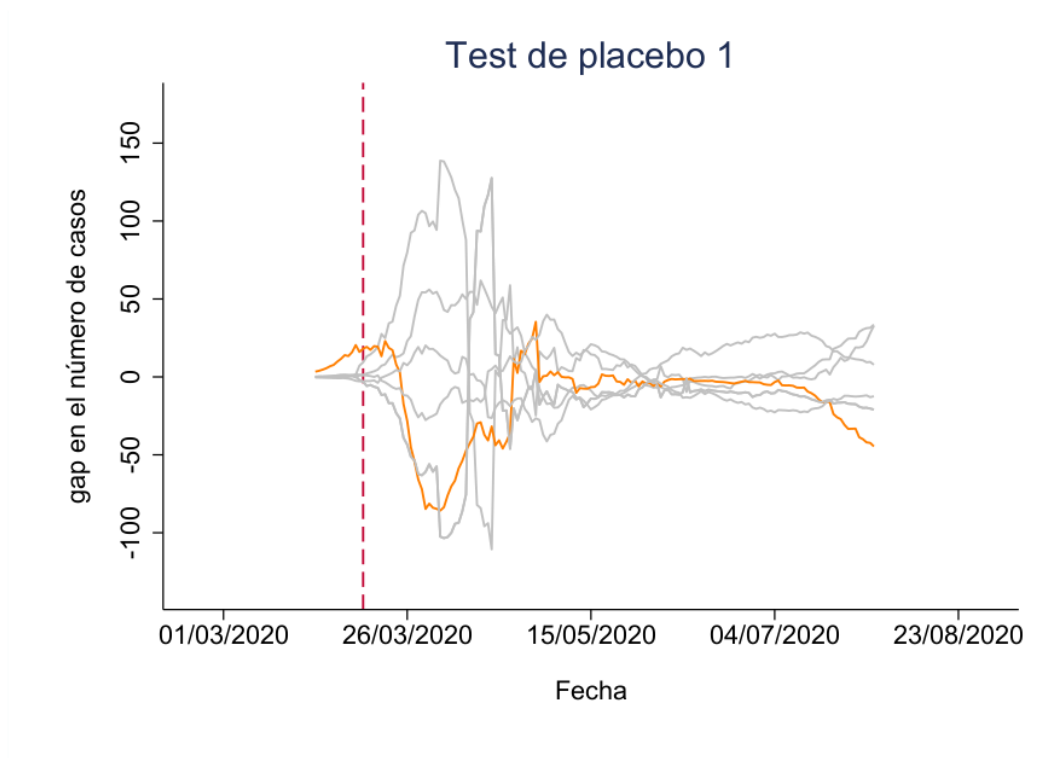


Figura 24: Brechas del numero de casos en España y brechas placebo para el resto de países

La **Figura 24** muestra los resultados del test placebo, las líneas grises muestran la diferencia en el número de contagios entre cada país del grupo de donantes y respectiva versión sintética. La línea naranja denota la brecha estima para España. Como se muestra en la figura, la brecha estimada para España es grande en relación con la distribución de brechas en el grupo donante. Por lo que podríamos decir que nuestro análisis proporciona una evidencia significativa.

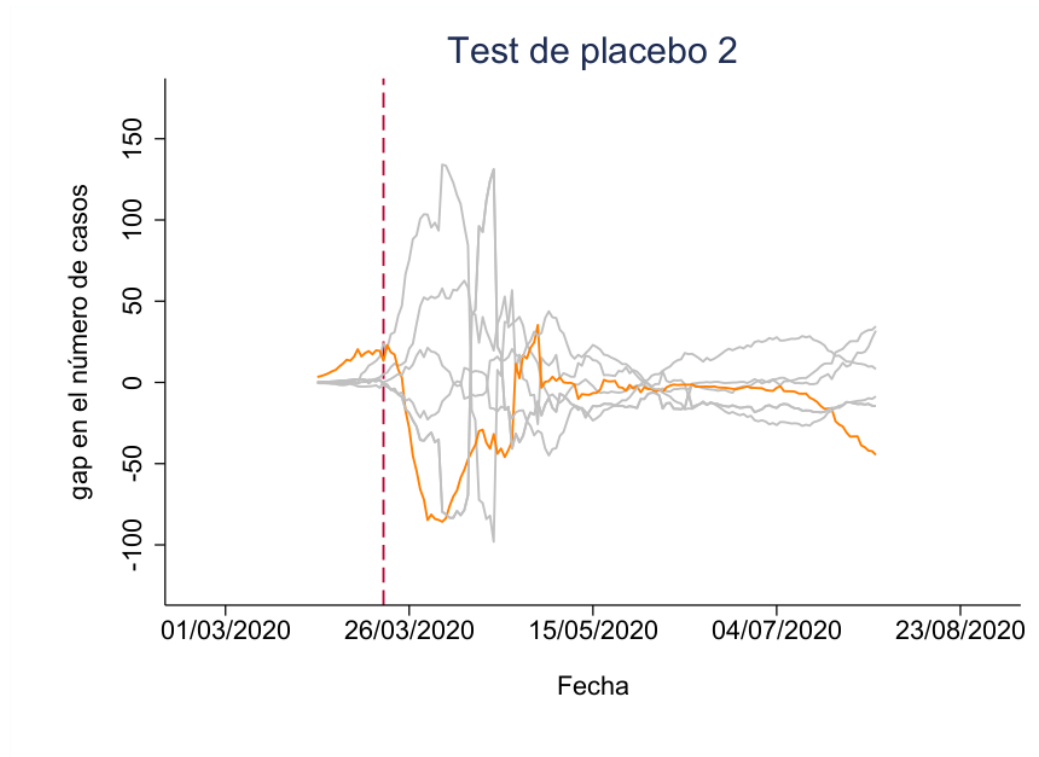


Figura 25: Brechas del numero de casos en España y brechas placebo para el resto de países

La **Figura 25** muestra el test de placebo realizado al segundo análisis propuesto en el trabajo como alternativa al primero. Como podemos apreciar la brecha estimada para España es grande en relación a la distribución del resto de brechas. Podemos concluir que este segundo análisis también proporciona unas estimaciones significativas.